



Adobe® Acrobat® Capture Plug-In

Online Guide

- Commands
- Using the Capture plug-in
- Capturing documents
- Correcting captured documents
- Setting Capture plug-in preferences
- Troubleshooting



Commands

File

Preferences



Capture...

Document

Capture Pages...

Edit

Show Capture Suspects

Find First Suspect



Using the Capture plug-in

You use the Capture plug-in when you choose the Capture Pages command in Acrobat Exchange. The plug-in uses optical character recognition (OCR) to convert bitmap text to text that can be corrected, indexed, searched, or copied to other files. The text it converts is in PDF image documents that were scanned directly, or imported, into Acrobat. For information about these PDF documents, and about the kinds of PDF documents produced by capturing text from them, see [About PDF document styles](#).

You can configure the Capture plug-in to recognize any of eight languages, hide recognized text behind a document image, and downsample images to minimize file size. For more information, see [Setting Capture plug-in preferences](#).

After you capture a document, you can use the touch-up tool in Exchange to review and correct text. See [Correcting captured documents](#) for details. ➤



The Capture plug-in is installed from the Acrobat 3.0 CD-ROM. For details on running the Acrobat Installer, see the *Acrobat Getting Started Guide*.

Use the Capture plug-in to convert small collections of paper documents and electronic images to PDF directly from Acrobat Exchange. If you need to convert large collections of paper documents or electronic images to PDF, consider upgrading to the full Acrobat Capture product, which offers a number of automated processing features and enhanced reviewing capabilities. ■



About PDF document styles

Adobe Acrobat can produce three styles of PDF documents:

- **PDF Image Only** documents contain only a bitmap picture of the original document. PDF Image Only files are produced by the Scan and the Import commands in Exchange. If you need to convert and view image files quickly, PDF Image Only is sufficient.
- **PDF Normal** documents contain electronic text that is scalable and can be indexed, searched, and copied. Page formatting and graphical images are preserved. You can create this kind of file with Acrobat Distiller, PDF Writer, or the Capture Pages command in Exchange. PDF Normal files are significantly smaller than their PDF Image Only counterparts, making them ideal for online distribution. ➤



■ **PDF Original Image with Hidden Text** documents combine features of PDF Image Only and PDF Normal documents. They contain a complete bitmap picture of the original document, but with recognized text hidden behind the picture. This provides the advantages of searchable text while ensuring that a document is identical in appearance to the original. Use this kind of PDF file when you are required to keep the original scanned image of a document for legal or archival purposes. PDF Original Image with Hidden Text files can be created only with the Capture Pages command in Exchange.

When you capture a PDF Image Only file, it usually reduces file size significantly. In most cases, PDF files captured with the PDF Normal setting are smaller than those captured with the PDF Original Image with Hidden Text setting. For details, see [Comparing PDF file sizes](#). ■



Capturing documents

The documents you capture are image files that you have scanned or imported into Exchange—**PDF Image Only** files.

To capture a document:

- 1 Using Exchange, do one of the following:
 - To import an image file, choose File > Import > Image. Select the file you want to import and click Open. See **Acrobat Exchange** for more information on importing image files.
 - To scan a paper document, choose File > Scan. Choose a scanner device and document type; then click Scan. See **Acrobat Scan** for information on using a scanner with Acrobat Exchange.
- 2 Choose Document > Capture Pages.
- 3 Determine which pages you want to capture by selecting All Pages, Current Page, or Specified Range and entering the page numbers in the text box. ➤



4 If you want to change the Capture preferences, click Preferences. The document will be captured with the new settings. See **Setting Capture plug-in preferences** for details.

5 Click OK. The Capture progress window shows the page, character, and word recognition process. In order for the process to be successful, the resolution of the captured PDF Image Only file must fall within the following ranges:

- Monochrome images, 200–600 dpi
- Grayscale or color images, 200–400 dpi

Also, the text should be dark against a light background. Text on a dark or shaded background, or on a page with complex color gradients, may not be recognized. ■



Correcting captured documents

When the Capture plug-in suspects it has not recognized a word correctly, it displays the bitmap image of the original word in the document and hides its best guess for the word behind the bitmap. This ensures accurate reproduction of the original, even without correction.

You can review and correct *suspect* words in Exchange with the touch-up tool. This is useful when you want your document to be fully searchable, for example, when indexing it for publication on CD-ROM or the World Wide Web.

Note: *The Capture plug-in uses the current PDF Writer settings for font embedding and subsetting when it creates PDF files. To avoid problems when correcting a captured document, be sure that font subsetting is not selected in PDF Writer before capturing the document. See [PDF Writer](#) for more information.* ➤



To view all the suspect words in a document:

Choose Edit > Show Capture Suspects. Each of the suspect words in the document is highlighted.

To review and correct suspect words:

1 Choose Edit > Find First Suspect. The suspect text is highlighted, and its bitmap image appears in the Suspect Image window.

2 Choose from one of the following:

- To accept the highlighted suspect text as correct, click Accept (TAB). The bitmap image is discarded.
- To leave the bitmap image in place, click Next (Shift+TAB).
- To edit the highlighted suspect text, type the correct text; then click Accept (TAB). The text is changed, and the bitmap image is discarded.

See [Acrobat Exchange](#) for more information on using the touch-up tool. ■



Setting Capture plug-in preferences

Choose File > Preferences > Capture to control the following Capture plug-in preferences:

- **Primary OCR Language** indicates which language dictionary is used to recognize words when documents are captured. The Capture plug-in also uses a custom dictionary, which you can modify. For more information, see [Adding words to the custom dictionary](#).
- **PDF Output Style** specifies what kind of PDF document the Capture plug-in creates. Two options are available: PDF Normal and PDF Original Image with Hidden Text. See [About PDF document styles](#) for details. ➤



- **Downsample Images** gives you the option of down-sampling images in captured PDF documents, which can be useful if you want to minimize file size. See [Choosing downsampling options](#) for details.
- **Location for Temporary Files** specifies the directory where temporary files are stored during the capture process. If you are running out of space in the specified directory, you can change the location by typing a new directory path. ■



Adding words to the custom dictionary

In addition to its standard language dictionaries, the Capture plug-in uses a custom dictionary to recognize words. You can add words to this custom dictionary by editing the dictionary file.

To edit the custom dictionary (Windows):

- 1** Using a text editor such as Notepad, or a word processor, open the custdict.spl file in the Acrobat3\Capture directory. If you use a word processor, open the file as a text file.
- 2** On a separate line, type each word you want to add. Be sure that the list remains in alphabetical order.
- 3** Save the file. If you are using a word processor, save the file as a text file.



Choosing downsampling options

When you capture a document, you can choose to downsample images in the captured document. Doing this can significantly reduce file size. With downsampling on, images are downsampled as follows:

- Black-and-white images are downsampled to 200dpi (Image less than 300 dpi are not downsampled)
- Grayscale and color images are downsampled to 150 dpi (Images less than 225 dpi are not downsampled)

Note: *When both the PDF Original Image + Text and downsampling options are selected, color and grayscale page images are downsampled below 200 dpi. Consequently, you will not be able to reprocess the resulting files with Capture. ➤*



To turn downsampling on or off:

- 1 Choose File > Preferences > Capture or Document > Capture Pages, and click the Preferences button.
- 2 Select or deselect the Downsample Images option as desired. (Downsampling is on by default.)

Note: *The Downsampling Images setting in Capture Preferences overrides the downsampling setting in PDF Writer.* ■



Comparing PDF file sizes

Consider file size when planning work flow or publishing documents online. Larger files (especially 24-bit color ones) take more time to capture, send over networks, and display on-screen.

The two charts on the following pages show the file sizes resulting from the import and capture of an 8 1/2-by-11 inch page containing text, line art, and a photograph. These charts show how the PDF Output Style you choose can effect the size of the final PDF file.

Note: *The Capture plug-in uses PDF Writer to create PDF files, but you cannot change PDF Writer compression settings to reduce the size of captured files. The Capture plug-in always uses the default PDF Writer compression settings. ➤*



Adeline & Associates



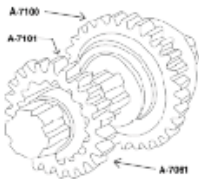
VINTAGE AUTOMOBILES

Automotive Professionals with Thirty Years Experience

Welcome to the world of Adeline and Associates. We offer complete or partial show-quality restoration, whether we do it for you or you do it yourself. At Adeline's, a full complement of eight craftspeople bring both broad knowledge and specialized skills to the task, from simple repair to rebuilding your engine to the original factory specs. Come to our huge showroom or check out our list of current inventory. We offer low-interest financing, and a unique 1-year buy-back guarantee. If you see something you like, we'll give you time to think about it: a 25% fully refundable deposit will hold any car for one month.



Photo courtesy of John Martin



Our giant selection of parts and supplies for interiors and exteriors, the largest on the West Coast—both originals and reproductions—simply has to be seen to be believed!

- ★ Windshields, running board mats, tail lights, mirrors
- ★ Suspension kits, brakes, wiring harnesses
- ★ Trim, emblems, convertible tops

Example file used for charts on following two pages ➤



Captured file sizes: downsampling on

In most cases, the size of a captured file is significantly reduced when downsampling is turned on in the Capture preferences.

Image type Scanned at 300 dpi	TIFF image file (original file)	PDF Image Only	PDF Normal	PDF Original Image + Hidden Text
	uncompressed ----- compressed			
Black and white	1043K ----- LZW 190K Group 4 201K	202K	61K	94K
4-bit grayscale	8335K ----- LZW 606K	421K	222K	325K
8-bit grayscale	8333K ----- LZW 1343K	1183K	113K	309K
8-bit indexed color	8335K ----- LZW 796K	612K	364K	477K
24-bit RGB color	24998K ----- LZW 3085K	2531K	500K	616K



Captured file sizes: downsampling off

The file in this chart was processed with downsampling turned off in the Capture preferences. (Turning downsampling on could reduce the file size even more). In some cases, a PDF Original Image + Hidden Text file is smaller than the PDF Image only file because the Capture plug-in uses additional compression methods.

Image type Scanned at 300 dpi	TIFF image file	PDF Image Only	PDF Normal	PDF Original Image + Hidden Text
	uncompressed ----- compressed			
Black and white	1043K ----- LZW 190K Group 4 201K	202K	161K	213K
4-bit grayscale	8335K ----- LZW 606K	421K	222K	328K
8-bit grayscale	8333K ----- LZW 1343K	1183K	766K	884K
8-bit indexed color	8335K ----- LZW 796K	612K	366K	477K
24-bit RGB color	24998K ----- LZW 3085K	2531K	1832K	2044K

© 1996 Adobe Systems Incorporated. All rights reserved.

Adobe Acrobat 3.0 Capture Online Guide

This manual, as well as the software described in it, is furnished under license and may be used or copied only in accordance with the terms of such license. The content of this manual is furnished for informational use only, is subject to change without notice, and should not be construed as a commitment by Adobe Systems Incorporated. Adobe Systems Incorporated assumes no responsibility or liability for any errors or inaccuracies that may appear in this book.

The copyrighted software that accompanies this manual is licensed to the End User for use only in strict accordance with the End User License Agreement, which the Licensee should read carefully before commencing use of the software. Except as permitted by such license, no part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, recording, or otherwise, without the prior written permission of Adobe Systems Incorporated.

Adobe, the Adobe logo, Acrobat, Acrobat Capture, the Acrobat logo, Distiller, Acrobat Exchange, Adobe Type Manager, PostScript, and the tagline "If you can dream it, you can do it" are trademarks of Adobe Systems Incorporated. Microsoft and Windows are registered trademarks and ActiveX and Windows NT are trademarks of Microsoft Corporation in the U.S. and other countries. Apple, Macintosh, Power Macintosh, and QuickTime are registered trademarks and AppleScript and TrueType are trademarks of Apple Computer, Inc. Lotus Notes is a registered trademark of Lotus Development Corporation. Netscape and Netscape Navigator are trademarks of Netscape Communications Corporation. UNIX is a registered trademark in the U.S. and other countries, licensed exclusively through X/Open Company, Ltd. Pentium is a trademark of Intel Corporation. All other products or name brands are trademarks of their respective owners.

This product contains an implementation of the LZW algorithm licensed under U.S. Patent 4,558,302. ➤

This software includes software licensed from Verity, Inc., copyright 1994. All rights reserved. The address of Verity, Inc., is 894 Ross Drive, Sunnyvale, California 94089. Verity ® and TOPIC ® are registered trademarks of Verity, Inc. in the United States and other countries. English Electronic Thesaurus copyright 1993 by INSO Corporation. Adapted from the Oxford Thesaurus copyright 1991 by Oxford University Press and from Roget's II: The New Thesaurus copyright 1980 by Houghton Mifflin Company. All rights reserved. Reproduction or disassembly of embodied programs and databases prohibited.

©1994 This software includes software licensed from RSA Data Security, Inc.

Written and designed at Adobe Systems Incorporated, 345 Park Ave, San Jose, CA 95110-2704.

Adobe Systems Europe Limited, Adobe House, 5 Mid New Cultins, Edinburgh EH11 4DU, Scotland, United Kingdom

Adobe Systems Co., Ltd., Yebisu Garden Place Tower, 4-20-3 Ebisu, Shibuya-ku, Tokyo 150, Japan

For defense agencies: Restricted Rights Legend. Use, reproduction, or disclosure is subject to restrictions set forth in subparagraph (c)(1)(ii) of the Rights in Technical Data and Computer Software clause at 252.227-7013.

For civilian agencies: Restricted Rights Legend. Use, reproduction, or disclosure is subject to restrictions set forth in subparagraphs (a) through (d) of the commercial Computer Software Restricted Rights clause at 52.227-19 and the limitations set forth in Adobe's standard commercial agreement for this software. Unpublished rights reserved under the copyright laws of the United States.

(9/96)



How to use this online guide



Page back or page forward.



Undoes a change of page or view, or redoes a change (Go Back/Go Forward).



Go to the Contents.



Go to the Index.



Go to the how-to page (this page).



Go to the “parent” of the current topic.

text

Go to the indicated topic.



Go to the next page of a continued topic.



End of a continued topic.

For instructions on printing this guide, go to the next page.



How to print this online guide

You can print separate topics or the entire guide. Since the pages of the guide have been made small for online viewing, Windows and Macintosh users may prefer to print them two to a page of paper—"two up."

To print pages two up:

1 Choose File > Print Setup (Windows) or File > Page Setup (Macintosh).

2 Follow the instructions for your platform:

- In Windows, click Options, select 2 up on the Paper tab, click OK to return to the Print Setup dialog box, and click OK again to close it. ➤

- On a Macintosh, choose 2 Up from the Layout menu and click OK.

Note: *If you can't perform step 2, you may not be using an Adobe or PostScript printer driver. If you are and you still can't perform the step, install the Adobe printer driver on the Acrobat CD-ROM. See the Acrobat Getting Started guide for installation instructions.*

3 Choose File > Print.

4 Indicate the page range.

Click OK (Windows) or Print (Macintosh). ■